Linear Algebra

Lecture 10 - Scalar Product

Oskar Kędzierski

11 December 2023

Scalar Product

Definition

A (standard) scalar product of two vectors $v = (v_1, \ldots, v_n), w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ is the real number

$$v \cdot w = \sum_{i=1}^{n} v_i w_i.$$

Scalar Product

Definition

A (standard) scalar product of two vectors $v = (v_1, \ldots, v_n), w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ is the real number

$$v \cdot w = \sum_{i=1}^{n} v_i w_i.$$

Example

Let $v = (1, 0, -2, 3), w = (0, 2, 2, 1) \in \mathbb{R}^4$. Then $v \cdot w = 1 \cdot 0 + 0 \cdot 2 - 2 \cdot 2 + 3 \cdot 1 = -1$.

Properties of Scalar Product

Let $v, v', w, w' \in \mathbb{R}^n$ and let $\alpha \in \mathbb{R}$. Then

Properties of Scalar Product

Let $v, v', w, w' \in \mathbb{R}^n$ and let $\alpha \in \mathbb{R}$. Then

i) $v \cdot w = w \cdot v$,

Properties of Scalar Product

Let $v, v', w, w' \in \mathbb{R}^n$ and let $\alpha \in \mathbb{R}$. Then

i) $v \cdot w = w \cdot v$,

ii) $(\alpha v) \cdot w = \alpha(v \cdot w)$,

Properties of Scalar Product

Let $v, v', w, w' \in \mathbb{R}^n$ and let $\alpha \in \mathbb{R}$. Then

i) $v \cdot w = w \cdot v$,

ii) $(\alpha v) \cdot w = \alpha(v \cdot w)$,

iii) $(v + v') \cdot w = v \cdot w + v' \cdot w$, $v \cdot (w + w') = v \cdot w + v \cdot w'$,

Properties of Scalar Product

Let $v, v', w, w' \in \mathbb{R}^n$ and let $\alpha \in \mathbb{R}$. Then

i) $v \cdot w = w \cdot v$,

ii) $(\alpha v) \cdot w = \alpha(v \cdot w)$,

iii) $(v + v') \cdot w = v \cdot w + v' \cdot w$, $v \cdot (w + w') = v \cdot w + v \cdot w'$,

iv) $v \cdot v > 0$ for $v \neq 0$.

Length of a Vector

Definition

The length of a vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ is the number

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}.$$

Length of a Vector

Definition

The length of a vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ is the number

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}.$$

Obviously $\|v\| \geqslant 0$ and

$$\|v\| = 0 \iff v = 0.$$

Note that if $\alpha \in \mathbb{R}$ then $\|\alpha v\| = |\alpha| \|v\|$.

Length of a Vector

Definition

The length of a vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ is the number

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}.$$

Obviously $\|v\| \geqslant 0$ and

$$\|v\| = 0 \iff v = 0.$$

Note that if $\alpha \in \mathbb{R}$ then $\|\alpha v\| = |\alpha|\|v\|$. In particular, if $v \neq 0$ then $\left\|\frac{v}{\|v\|}\right\| = 1$. The vector $\frac{v}{\|v\|}$ is called the **normalized vector** of v.

Length of a Vector

Definition

The length of a vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ is the number

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}.$$

Obviously $\|v\| \geqslant 0$ and

$$\|v\| = 0 \iff v = 0.$$

Note that if $\alpha \in \mathbb{R}$ then $\|\alpha v\| = |\alpha|\|v\|$. In particular, if $v \neq 0$ then $\left\|\frac{v}{\|v\|}\right\| = 1$. The vector $\frac{v}{\|v\|}$ is called the **normalized vector** of v.

Definition

Two vectors $v, w \in \mathbb{R}^n$ are said to be **orthogonal** (or perpendicular) if $v \cdot w = 0$. We write $v \perp w$.

Geometric Interpretation and the Law of Cosines

Proposition

For any $v, w \in \mathbb{R}^n$

$$v \cdot w = \|v\|\|w\| \cos \angle(v, w),$$

where $\angle(v, w)$ *is the angle between* v *and* w.

Geometric Interpretation and the Law of Cosines

Proposition

For any $v, w \in \mathbb{R}^n$

$$v \cdot w = \|v\|\|w\| \cos \angle(v, w),$$

where $\angle(v, w)$ *is the angle between* v *and* w.

Proof.

From the law of cosines

$$\|v - w\|^2 = (v - w) \cdot (v - w) = \|v\|^2 + \|w\|^2 - 2(v \cdot w) =$$

$$= \|v\|^2 + \|w\|^2 - 2\|v\|\|w\| \cos \angle(v, w).$$

$\square$

Cauchy–Schwarz Inequality

Corollary (Cauchy-Schwarz inequality)

For any $v, w \in \mathbb{R}^n$

$$\|v\|\|w\| \geqslant |v \cdot w|,$$

and the equality holds if and only if v, w *are linearly dependent.*

Cauchy–Schwarz Inequality

Corollary (Cauchy-Schwarz inequality)

For any $v, w \in \mathbb{R}^n$

$$\|v\|\|w\| \geqslant |v \cdot w|,$$

and the equality holds if and only if v, w *are linearly dependent.*

Proof.

In general

$$|\cos \angle(v, w)| \leqslant 1,$$

and

$$|\cos \angle(v, w)| = 1 \iff \angle(v, w) \in \{0, \pi\}.$$

$\square$

Pythagorean Theorem

Example

Let $v = (3, 0, 4), w = (0, 1, 0), u = (1, 1, 1)$. Then $\|v\| = \sqrt{3^2 + 0^2 + 4^2} = \sqrt{9 + 16} = 5$. The normalized vector of v is $\frac{1}{5}(3, 0, 4)$. Since $v \cdot w = 3 \cdot 0 + 0 \cdot 1 + 4 \cdot 0 = 0$ then $v \perp w$ but w is not orthogonal to u because $w \cdot u = 1$.

Pythagorean Theorem

Example

Let $v = (3, 0, 4), w = (0, 1, 0), u = (1, 1, 1)$. Then $\|v\| = \sqrt{3^2 + 0^2 + 4^2} = \sqrt{9 + 16} = 5$. The normalized vector of v is $\frac{1}{5}(3, 0, 4)$. Since $v \cdot w = 3 \cdot 0 + 0 \cdot 1 + 4 \cdot 0 = 0$ then $v \perp w$ but w is not orthogonal to u because $w \cdot u = 1$.

Theorem (Pythagoras)

If $v \perp w$ then $\|v + w\|^2 = \|v\|^2 + \|w\|^2$.

Pythagorean Theorem

Example

Let $v = (3, 0, 4), w = (0, 1, 0), u = (1, 1, 1)$. Then $\|v\| = \sqrt{3^2 + 0^2 + 4^2} = \sqrt{9 + 16} = 5$. The normalized vector of v is $\frac{1}{5}(3, 0, 4)$. Since $v \cdot w = 3 \cdot 0 + 0 \cdot 1 + 4 \cdot 0 = 0$ then $v \perp w$ but w is not orthogonal to u because $w \cdot u = 1$.

Theorem (Pythagoras)

If $v \perp w$ then $\|v + w\|^2 = \|v\|^2 + \|w\|^2$.

Proof.

$\|v + w\|^2 = (v + w) \cdot (v + w) = v \cdot v + v \cdot w + w \cdot v + w \cdot w = \|v\|^2 + \|w\|^2$. $\square$

Orthogonal Complement

Let $A \subset \mathbb{R}^n$ be any set. Let

$$A^{\perp} = \{w \in \mathbb{R}^n \mid w \cdot v = 0 \text{ for all } v \in A\}.$$

Orthogonal Complement

Let $A \subset \mathbb{R}^n$ be any set. Let

$$A^\perp = \{w \in \mathbb{R}^n \mid w \cdot v = 0 \text{ for all } v \in A\}.$$

The set $A^\perp$ is a subspace of $\mathbb{R}^n$.

Orthogonal Complement

Let $A \subset \mathbb{R}^n$ be any set. Let

$$A^\perp = \{w \in \mathbb{R}^n \mid w \cdot v = 0 \text{ for all } v \in A\}.$$

The set $A^\perp$ is a subspace of $\mathbb{R}^n$.

Definition

Let $V \subset \mathbb{R}^n$ be a subspace. The **orthogonal complement** of V in $\mathbb{R}^n$ is $V^\perp$.

Orthogonal Complement

Let $A \subset \mathbb{R}^n$ be any set. Let

$$A^\perp = \{w \in \mathbb{R}^n \mid w \cdot v = 0 \text{ for all } v \in A\}.$$

The set $A^\perp$ is a subspace of $\mathbb{R}^n$.

Definition

Let $V \subset \mathbb{R}^n$ be a subspace. The **orthogonal complement** of V in $\mathbb{R}^n$ is $V^\perp$.

Example

Let $V = \text{lin}((1,2)) \subset \mathbb{R}^2$. Then $V^\perp = \text{lin}((2,-1))$.

Properties

Proposition

Let $v_1, \ldots, v_k \in \mathbb{R}^n$. *Then*

$$(\mathrm{lin}(v_1, \ldots, v_k))^\perp = \{v_1, \ldots, v_k\}^\perp.$$

Properties

Proposition

Let $v_1, \ldots, v_k \in \mathbb{R}^n$. *Then*

$$(\mathrm{lin}(v_1, \ldots, v_k))^\perp = \{v_1, \ldots, v_k\}^\perp.$$

Proof.

Set $V = \mathrm{lin}(v_1, \ldots, v_k)$. Assume $w \in V^\perp$. Then, in particular, $w \cdot v_i = 0$, hence $V^\perp \subset \{v_1, \ldots, v_k\}^\perp$.

Properties

Proposition

Let $v_1, \ldots, v_k \in \mathbb{R}^n$. *Then*

$$(\mathrm{lin}(v_1, \ldots, v_k))^\perp = \{v_1, \ldots, v_k\}^\perp.$$

Proof.

Set $V = \mathrm{lin}(v_1, \ldots, v_k)$. Assume $w \in V^\perp$. Then, in particular, $w \cdot v_i = 0$, hence $V^\perp \subset \{v_1, \ldots, v_k\}^\perp$. If $w \cdot v_i = 0$ for $i = 1, \ldots, k$ then for any $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, k$

$$w \cdot (\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k) = \alpha_1 (w \cdot v_1) + \alpha_2 (w \cdot v_2) + \ldots + \alpha_k (w \cdot v_k) = 0.$$

$\square$

Example

Let

$$V = \mathrm{lin}((1,2,3,1),(1,3,2,2),(2,5,5,3)) \subset \mathbb{R}^4.$$

Then

$$V^{\perp}: \begin{cases} x_1 + 2x_2 + 3x_3 + x_4 = 0 \\ x_1 + 3x_2 + 2x_3 + 2x_4 = 0 \\ 2x_1 + 5x_2 + 5x_3 + 3x_4 = 0 \end{cases}$$

The solution of that system is equal to

$$\begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 3 & 2 & 2 \\ 2 & 5 & 5 & 3 \end{bmatrix} \longrightarrow \ldots \longrightarrow \begin{bmatrix} 1 & 0 & 5 & -1 \\ 0 & 1 & -1 & 1 \end{bmatrix}$$

that is,

$$V^{\perp} = \mathrm{lin}((-5,1,1,0),(1,-1,0,1)).$$

Properties (continued)

Proposition

Let $V \subset \mathbb{R}^n$, $\dim V = k$. *Then* $\dim V^\perp = n - k$ *and* $V \cap V^\perp = \{0\}$.

Properties (continued)

Proposition

Let $V \subset \mathbb{R}^n$, $\dim V = k$. *Then* $\dim V^\perp = n - k$ *and* $V \cap V^\perp = \{0\}$.

Proof.

Let $v_1, \dots v_k$ be a basis of V, where $v_i = (a_{i1}, a_{i2}, \dots, a_{in})$. By the above Proposition $(x_1, \dots, x_n) \in V^\perp$ if and only if it is a solution of the system of linear equations

$$V^\perp : \begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = 0 \\ \vdots & & \vdots & & \ddots & & \vdots & \vdots \\ a_{k1}x_1 & + & a_{k2}x_2 & + & \dots & + & a_{kn}x_n & = 0 \end{cases}$$

$\square$

Properties (continued)

Proof.

The rank of the matrix $\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kn} \end{bmatrix}$ is equal to k, hence by the Kronecker-Capelli theorem the dimension of the set of solutions is $n - k$.

Properties (continued)

Proof.

The rank of the matrix $\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kn} \end{bmatrix}$ is equal to k, hence by the Kronecker-Capelli theorem the dimension of the set of solutions is $n - k$. Moreover, if $w \in V \cap V^{\perp}$ then $w \cdot w = 0$ hence $w = 0$. $\square$

Properties (continued)

Proof.

The rank of the matrix $\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kn} \end{bmatrix}$ is equal to k, hence by the Kronecker-Capelli theorem the dimension of the set of solutions is $n-k$. Moreover, if $w \in V \cap V^{\perp}$ then $w \cdot w = 0$ hence $w = 0$. $\square$

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then $(V^{\perp})^{\perp} = V$.

Properties (continued)

Proof.

The rank of the matrix $\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kn} \end{bmatrix}$ is equal to k, hence by the Kronecker-Capelli theorem the dimension of the set of solutions is $n-k$. Moreover, if $w \in V \cap V^{\perp}$ then $w \cdot w = 0$ hence $w = 0$. $\square$

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then $(V^{\perp})^{\perp} = V$.

Proof.

By the above $\dim(V^{\perp})^{\perp} = n - \dim V^{\perp} = n - (n - \dim V)$. Since $V \subset (V^{\perp})^{\perp}$ and both have the same dimension they are equal. $\square$

Example

Let $V \subset \mathbb{R}^2$ be subspace given by the linear equation $2x_1 + 3x_2 = 0$. Then $V = \mathrm{lin}((-3, 2))$ and $V^\perp = \mathrm{lin}((2, 3))$.

Example

Let $V \subset \mathbb{R}^2$ be subspace given by the linear equation $2x_1 + 3x_2 = 0$. Then $V = \mathrm{lin}((-3, 2))$ and $V^\perp = \mathrm{lin}((2, 3))$.

This can be generalized to

Proposition

Let $V \subset \mathbb{R}^n$ be equal to the set of solutions of the system of linear equations

$$V: \begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = 0 \\ \vdots & & \vdots & & \ddots & & \vdots & \vdots \\ a_{k1}x_1 & + & a_{k2}x_2 & + & \dots & + & a_{kn}x_n & = 0 \end{cases}$$

Example

Let $V \subset \mathbb{R}^2$ be subspace given by the linear equation $2x_1 + 3x_2 = 0$. Then $V = \mathrm{lin}((-3, 2))$ and $V^\perp = \mathrm{lin}((2, 3))$.

This can be generalized to

Proposition

Let $V \subset \mathbb{R}^n$ be equal to the set of solutions of the system of linear equations

$$V: \begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = 0 \\ \vdots & & \vdots & & \ddots & & \vdots & \vdots \\ a_{k1}x_1 & + & a_{k2}x_2 & + & \dots & + & a_{kn}x_n & = 0 \end{cases}$$

Then

$$V^\perp = \mathrm{lin}((a_{11}, a_{12}, \dots, a_{1n}), \dots, (a_{k1}, a_{k2}, \dots, a_{kn})).$$

Proof.

Let $v_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ for $i = 1, \ldots, k$. Then

$$V = \{v_1, v_2, \ldots, v_k\}^{\perp}.$$

Proof.

Let $v_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ for $i = 1, \ldots, k$. Then

$$V = \{v_1, v_2, \ldots, v_k\}^{\perp}.$$

Hence

$$V^{\perp} = (\{v_1, v_2, \ldots, v_k\}^{\perp})^{\perp} = ((\mathrm{lin}(v_1, v_2, \ldots, v_k))^{\perp})^{\perp} =$$

$$= \mathrm{lin}(v_1, v_2, \ldots, v_k).$$

$\square$

Proof.

Let $v_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ for $i = 1, \ldots, k$. Then

$$V = \{v_1, v_2, \ldots, v_k\}^{\perp}.$$

Hence

$$V^{\perp} = (\{v_1, v_2, \ldots, v_k\}^{\perp})^{\perp} = ((\mathrm{lin}(v_1, v_2, \ldots, v_k))^{\perp})^{\perp} =$$

$$= \mathrm{lin}(v_1, v_2, \ldots, v_k).$$

$\square$

Example

Let $V \subset \mathbb{R}^4$ be equal to the set of solutions of the system

$$\begin{cases} 2x_1 + 3x_2 + 4x_3 + 6x_4 = 0 \\ x_1 - 2x_2 + 5x_3 = 0 \end{cases}$$

Proof.

Let $v_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ for $i = 1, \ldots, k$. Then

$$V = \{v_1, v_2, \ldots, v_k\}^{\perp}.$$

Hence

$$V^{\perp} = (\{v_1, v_2, \ldots, v_k\}^{\perp})^{\perp} = ((\mathrm{lin}(v_1, v_2, \ldots, v_k))^{\perp})^{\perp} =$$

$$= \mathrm{lin}(v_1, v_2, \ldots, v_k).$$

$\square$

Example

Let $V \subset \mathbb{R}^4$ be equal to the set of solutions of the system

$$\begin{cases} 2x_1 + 3x_2 + 4x_3 + 6x_4 = 0 \\ x_1 - 2x_2 + 5x_3 = 0 \end{cases}$$

Then $V^{\perp} = \mathrm{lin}((2, 3, 4, 6), (1, -2, 5, 0))$.

Orthogonal Basis

Let $V \subset \mathbb{R}^n$ be a subspace of $\mathbb{R}^n$.

Definition

Let $\mathcal{A} = (v_1, \ldots, v_k)$ be a basis of subspace V. The basis $\mathcal{A}$ is said to be **orthogonal** if $v_i \perp v_j$ for $i \neq j$ and $i, j = 1, \ldots, k$. The basis $\mathcal{A}$ is said to be **orthonormal** if it is orthogonal and $\|v_i\| = 1$ for $i = 1, \ldots, k$, i.e. each vector is of length 1.

Orthogonal Basis

Let $V \subset \mathbb{R}^n$ be a subspace of $\mathbb{R}^n$.

Definition

Let $\mathcal{A} = (v_1, \ldots, v_k)$ be a basis of subspace V. The basis $\mathcal{A}$ is said to be **orthogonal** if $v_i \perp v_j$ for $i \neq j$ and $i, j = 1, \ldots, k$. The basis $\mathcal{A}$ is said to be **orthonormal** if it is orthogonal and $\|v_i\| = 1$ for $i = 1, \ldots, k$, i.e. each vector is of length 1.

Examples

i) the standard basis $\varepsilon_1 = (1, 0, 0, \ldots, 0), \varepsilon_2 = (0, 1, 0, \ldots, 0), \ldots, \varepsilon_n = (0, 0, 0, \ldots, 1)$ of $\mathbb{R}^n$ is orthonormal,

Orthogonal Basis

Let $V \subset \mathbb{R}^n$ be a subspace of $\mathbb{R}^n$.

Definition

Let $\mathcal{A} = (v_1, \ldots, v_k)$ be a basis of subspace V. The basis $\mathcal{A}$ is said to be **orthogonal** if $v_i \perp v_j$ for $i \neq j$ and $i, j = 1, \ldots, k$. The basis $\mathcal{A}$ is said to be **orthonormal** if it is orthogonal and $\|v_i\| = 1$ for $i = 1, \ldots, k$, i.e. each vector is of length 1.

Examples

i) the standard basis $\varepsilon_1 = (1, 0, 0, \ldots, 0), \varepsilon_2 = (0, 1, 0, \ldots, 0), \ldots, \varepsilon_n = (0, 0, 0, \ldots, 1)$ of $\mathbb{R}^n$ is orthonormal,

ii) the basis $(-1, 2, 2), (2, -1, 2), (2, 2, -1)$ is an orthogonal basis of $\mathbb{R}^3$ (but not orthonormal),

Orthogonal Basis

Let $V \subset \mathbb{R}^n$ be a subspace of $\mathbb{R}^n$.

Definition

Let $\mathcal{A} = (v_1, \ldots, v_k)$ be a basis of subspace V. The basis $\mathcal{A}$ is said to be **orthogonal** if $v_i \perp v_j$ for $i \neq j$ and $i, j = 1, \ldots, k$. The basis $\mathcal{A}$ is said to be **orthonormal** if it is orthogonal and $\|v_i\| = 1$ for $i = 1, \ldots, k$, i.e. each vector is of length 1.

Examples

i) the standard basis $\varepsilon_1 = (1, 0, 0, \ldots, 0), \varepsilon_2 = (0, 1, 0, \ldots, 0), \ldots, \varepsilon_n = (0, 0, 0, \ldots, 1)$ of $\mathbb{R}^n$ is orthonormal,

ii) the basis $(-1, 2, 2), (2, -1, 2), (2, 2, -1)$ is an orthogonal basis of $\mathbb{R}^3$ (but not orthonormal),

iii) the basis $(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3}), (\frac{2}{3}, -\frac{1}{3}, \frac{2}{3}), (\frac{2}{3}, \frac{2}{3}, -\frac{1}{3})$ is an orthonormal basis of $\mathbb{R}^3$.

Coordinates Relative to Orthogonal Basis

Proposition

Let $v_1, \ldots, v_k$ be an orthogonal basis of the subspace $V \subset \mathbb{R}^n$. For any $v \in V$

$$v = \frac{v \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{v \cdot v_2}{v_2 \cdot v_2} v_2 + \ldots + \frac{v \cdot v_k}{v_k \cdot v_k} v_k.$$

Coordinates Relative to Orthogonal Basis

Proposition

Let $v_1, \ldots, v_k$ be an orthogonal basis of the subspace $V \subset \mathbb{R}^n$. For any $v \in V$

$$v = \frac{v \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{v \cdot v_2}{v_2 \cdot v_2} v_2 + \ldots + \frac{v \cdot v_k}{v_k \cdot v_k} v_k.$$

Proof.

There exist unique $\alpha_i \in \mathbb{R}$ such that $v = \alpha_1 v_1 + \ldots + \alpha_k v_k$. Therefore

$$v \cdot v_i = \alpha_1 (v_1 \cdot v_i) + \ldots + \alpha_i (v_i \cdot v_i) + \ldots + \alpha_k (v_k \cdot v_i) = \alpha_i (v_i \cdot v_i),$$

since $v_i \cdot v_j = 0$ for $i \neq j$. $\square$

Coordinates Relative to Orthogonal Basis

Proposition

Let $v_1, \ldots, v_k$ be an orthogonal basis of the subspace $V \subset \mathbb{R}^n$. For any $v \in V$

$$v = \frac{v \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{v \cdot v_2}{v_2 \cdot v_2} v_2 + \ldots + \frac{v \cdot v_k}{v_k \cdot v_k} v_k.$$

Proof.

There exist unique $\alpha_i \in \mathbb{R}$ such that $v = \alpha_1 v_1 + \ldots + \alpha_k v_k$. Therefore

$$v \cdot v_i = \alpha_1 (v_1 \cdot v_i) + \ldots + \alpha_i (v_i \cdot v_i) + \ldots + \alpha_k (v_k \cdot v_i) = \alpha_i (v_i \cdot v_i),$$

since $v_i \cdot v_j = 0$ for $i \neq j$. $\square$

Corollary

If vectors $v_1, \ldots, v_k \in \mathbb{R}^n$ are pairwise orthogonal and $v_i \neq 0$ for $i = 1, \ldots, k$ then they are linearly independent.

Existence of Orthogonal Basis

Example

The coordinates of the vector $(1,1,1)$ relative to the orthogonal basis $(-1,2,2),(2,-1,2),(2,2,-1)$ of $\mathbb{R}^3$ are $\frac{1}{3},\frac{1}{3},\frac{1}{3}$ since $\frac{(1,1,1)\cdot(-1,2,2)}{(-1,2,2)\cdot(-1,2,2)}=\frac{1}{3}$, $\frac{(1,1,1)\cdot(2,-1,2)}{(2,-1,2)\cdot(2,-1,2)}=\frac{1}{3}$, $\frac{(1,1,1)\cdot(2,2,-1)}{(2,2,-1)\cdot(2,2,-1)}=\frac{1}{3}$, i.e.

$$(1,1,1)=\frac{1}{3}(-1,2,2)+\frac{1}{3}(2,-1,2)+\frac{1}{3}(2,2,-1).$$

Existence of Orthogonal Basis

Example

The coordinates of the vector $(1,1,1)$ relative to the orthogonal basis $(-1,2,2),(2,-1,2),(2,2,-1)$ of $\mathbb{R}^3$ are $\frac{1}{3},\frac{1}{3},\frac{1}{3}$ since $\frac{(1,1,1)\cdot(-1,2,2)}{(-1,2,2)\cdot(-1,2,2)}=\frac{1}{3}$, $\frac{(1,1,1)\cdot(2,-1,2)}{(2,-1,2)\cdot(2,-1,2)}=\frac{1}{3}$, $\frac{(1,1,1)\cdot(2,2,-1)}{(2,2,-1)\cdot(2,2,-1)}=\frac{1}{3}$, i.e.

$$(1,1,1)=\frac{1}{3}(-1,2,2)+\frac{1}{3}(2,-1,2)+\frac{1}{3}(2,2,-1).$$

Proposition

Any subspace $V\subset\mathbb{R}^n$ has an orthogonal basis.

Existence of Orthogonal Basis

Example

The coordinates of the vector $(1,1,1)$ relative to the orthogonal basis $(-1,2,2),(2,-1,2),(2,2,-1)$ of $\mathbb{R}^3$ are $\frac{1}{3},\frac{1}{3},\frac{1}{3}$ since $\frac{(1,1,1)\cdot(-1,2,2)}{(-1,2,2)\cdot(-1,2,2)}=\frac{1}{3}$, $\frac{(1,1,1)\cdot(2,-1,2)}{(2,-1,2)\cdot(2,-1,2)}=\frac{1}{3}$, $\frac{(1,1,1)\cdot(2,2,-1)}{(2,2,-1)\cdot(2,2,-1)}=\frac{1}{3}$, i.e.

$$(1,1,1)=\frac{1}{3}(-1,2,2)+\frac{1}{3}(2,-1,2)+\frac{1}{3}(2,2,-1).$$

Proposition

Any subspace $V\subset\mathbb{R}^n$ has an orthogonal basis.

Proof.

A proof will be given later. $\square$

Example

Example

Let $V \subset \mathbb{R}^3$ be given by the equation $x_1 + x_2 + x_3 = 0$. We compute inductively an orthogonal basis of V by choosing vectors orthogonal to the previously chosen ones. Let $v_1 = (1, 0, -1)$. To find $v_2 \in V$ such that $v_1 \perp v_2$ solve

$$\begin{cases} x_1 & + & x_2 & + & x_3 & = 0 \\ x_1 & & & - & x_3 & = 0 \end{cases} \iff \begin{cases} 2x_1 & + & x_2 & & & = 0 \\ x_1 & & & - & x_3 & = 0 \end{cases}$$

Example

Example

Let $V \subset \mathbb{R}^3$ be given by the equation $x_1 + x_2 + x_3 = 0$. We compute inductively an orthogonal basis of V by choosing vectors orthogonal to the previously chosen ones. Let $v_1 = (1, 0, -1)$. To find $v_2 \in V$ such that $v_1 \perp v_2$ solve

$$\begin{cases} x_1 + x_2 + x_3 = 0 \\ x_1 - x_3 = 0 \end{cases} \iff \begin{cases} 2x_1 + x_2 = 0 \\ x_1 - x_3 = 0 \end{cases}$$

$\iff x_2 = -2x_1,\ x_3 = x_1$. For example $v_2 = (1, -2, 1)$. Since $\dim V = 2$ vectors v_1, v_2 form an orthogonal basis of V. By taking normalized vectors we get an orthonormal basis $\frac{1}{\sqrt{2}}(1, 0, -1), \frac{1}{\sqrt{6}}(1, -2, 1)$ of V.

Orthogonal Decomposition

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then any vector $w \in \mathbb{R}^n$ can be written uniquely as

$$w = v + v^{\perp} \text{ where } v \in V,\ v^{\perp} \in V^{\perp}.$$

Orthogonal Decomposition

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then any vector $w \in \mathbb{R}^n$ can be written uniquely as

$$w = v + v^{\perp} \text{ where } v \in V,\ v^{\perp} \in V^{\perp}.$$

Proof.

Let $v_1, \ldots, v_k$ be a basis of V and let $v_{k+1}, \ldots, v_n$ be a basis of $V^{\perp}$.

Orthogonal Decomposition

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then any vector $w \in \mathbb{R}^n$ can be written uniquely as

$$w = v + v^{\perp} \text{ where } v \in V,\ v^{\perp} \in V^{\perp}.$$

Proof.

Let $v_1, \ldots, v_k$ be a basis of V and let $v_{k+1}, \ldots, v_n$ be a basis of $V^{\perp}$. Then

$$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0 \iff \begin{cases} \alpha_1 v_1 + \ldots + \alpha_k v_k = 0 \\ \alpha_{k+1} v_{k+1} + \ldots + \alpha_n v_n = 0 \end{cases} \iff$$

$$\iff \alpha_1 = \ldots = \alpha_n = 0,$$

hence $\mathcal{B} = (v_1, \ldots, v_n)$ is a basis of $\mathbb{R}^n$.

Orthogonal Decomposition

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then any vector $w \in \mathbb{R}^n$ can be written uniquely as

$$w = v + v^{\perp} \text{ where } v \in V,\ v^{\perp} \in V^{\perp}.$$

Proof.

Let $v_1, \ldots, v_k$ be a basis of V and let $v_{k+1}, \ldots, v_n$ be a basis of $V^{\perp}$. Then

$$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0 \iff \begin{cases} \alpha_1 v_1 + \ldots + \alpha_k v_k = 0 \\ \alpha_{k+1} v_{k+1} + \ldots + \alpha_n v_n = 0 \end{cases} \iff$$

$$\iff \alpha_1 = \ldots = \alpha_n = 0,$$

hence $\mathcal{B} = (v_1, \ldots, v_n)$ is a basis of $\mathbb{R}^n$. This proves the existence of a decomposition.

Orthogonal Decomposition

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then any vector $w \in \mathbb{R}^n$ can be written uniquely as

$$w = v + v^{\perp} \text{ where } v \in V,\ v^{\perp} \in V^{\perp}.$$

Proof.

Let $v_1, \ldots, v_k$ be a basis of V and let $v_{k+1}, \ldots, v_n$ be a basis of $V^{\perp}$. Then

$$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0 \iff \begin{cases} \alpha_1 v_1 + \ldots + \alpha_k v_k = 0 \\ \alpha_{k+1} v_{k+1} + \ldots + \alpha_n v_n = 0 \end{cases} \iff$$

$$\iff \alpha_1 = \ldots = \alpha_n = 0,$$

hence $\mathcal{B} = (v_1, \ldots, v_n)$ is a basis of $\mathbb{R}^n$. This proves the existence of a decomposition.

Orthogonal Decomposition (continued)

Proof.

If

$$w = v + v^{\perp} = u + u^{\perp},$$

where $v, u \in V,\ v^{\perp}, u^{\perp} \in V^{\perp}$, then

$$v - u = u^{\perp} - v^{\perp} \in V \cap V^{\perp} = \{0\}.$$

Orthogonal Decomposition (continued)

Proof.

If

$$w = v + v^{\perp} = u + u^{\perp},$$

where $v, u \in V,\ v^{\perp}, u^{\perp} \in V^{\perp}$, then

$$v - u = u^{\perp} - v^{\perp} \in V \cap V^{\perp} = \{0\}.$$

Therefore

$$v = u,\ v^{\perp} = u^{\perp}.$$

$\square$

Orthogonal Projection and Reflection

Definition

For any subspace $V \subset \mathbb{R}^n$ and $w \in \mathbb{R}^n$ the function $P_V : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$P_V(w) = v, \text{ where } w = v + v^{\perp}, v \in V, \ v^{\perp} \in V^{\perp},$$

is a linear transformation called **the orthogonal projection** on the subspace V.

Orthogonal Projection and Reflection

Definition

For any subspace $V \subset \mathbb{R}^n$ and $w \in \mathbb{R}^n$ the function $P_V : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$P_V(w) = v, \text{ where } w = v + v^{\perp}, v \in V, \ v^{\perp} \in V^{\perp},$$

is a linear transformation called **the orthogonal projection** on the subspace V.

Note that with the above notation $P_{V^{\perp}}(w) = v^{\perp}$, that is $w = P_V(w) + P_{V^{\perp}}(w)$. The linearity of P_V follows from the uniqueness of the orthogonal decomposition.

Orthogonal Projection and Reflection (continued)

Definition

For any subspace $V \subset \mathbb{R}^n$ and $w \in \mathbb{R}^n$ the function $S_V : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$S_V(w) = v - v^{\perp}, \text{ where } w = v + v^{\perp}, v \in V, \ v^{\perp} \in V^{\perp},$$

is a linear transformation called **the orthogonal reflection** across the subspace V.

Orthogonal Projection and Reflection (continued)

Definition

For any subspace $V \subset \mathbb{R}^n$ and $w \in \mathbb{R}^n$ the function $S_V : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$S_V(w) = v - v^{\perp}, \text{ where } w = v + v^{\perp}, v \in V, \ v^{\perp} \in V^{\perp},$$

is a linear transformation called **the orthogonal reflection** across the subspace V.

Note that

$$S_V(w) = P_V(w) - P_{V^{\perp}}(w) = 2P_V(w) - w.$$

Properties

Example

Let $V = \mathrm{lin}(v)$. Then $P_V(w) = \frac{w \cdot v}{v \cdot v} v$.

Properties

Example

Let $V = \text{lin}(v)$. Then $P_V(w) = \frac{w \cdot v}{v \cdot v} v$.

Proposition

i) $P_V(w) \in V$ *and* $(P_v(w) = w \iff w \in V)$,

Properties

Example

Let $V = \mathrm{lin}(v)$. Then $P_V(w) = \frac{w \cdot v}{v \cdot v} v$.

Proposition

i) $P_V(w) \in V$ *and* $(P_v(w) = w \iff w \in V)$,

ii) *let* $d(w, V) = \min\{\|w - v\| \mid v \in V\}$ *be the distance between the vector* w *and the subspace* V. *Then* $P_V(w)$ *is the unique vector in* V *such that* $d(w, V) = \|w - P_V(w)\|$,

Properties

Example

Let $V = \mathrm{lin}(v)$. Then $P_V(w) = \frac{w \cdot v}{v \cdot v} v$.

Proposition

i) $P_V(w) \in V$ *and* $(P_v(w) = w \iff w \in V)$,

ii) *let* $d(w, V) = \min\{\|w - v\| \mid v \in V\}$ *be the distance between the vector* w *and the subspace* V. *Then* $P_V(w)$ *is the unique vector in* V *such that* $d(w, V) = \|w - P_V(w)\|$,

iii) *if* $v_1, \ldots, v_k$ *is an orthogonal basis of* V *then*

$$P_V(w) = \frac{w \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{w \cdot v_2}{v_2 \cdot v_2} v_2 + \ldots + \frac{w \cdot v_k}{v_k \cdot v_k} v_k.$$

Properties (continued)

Proof.

ii) recall $w = P_V(w) + P_{V^\perp}(w)$, then for any $v \in V$, by the Pythagorean theorem $\|w - v\|^2 = \|(P_V(w) - v) + P_{V^\perp}(w)\|^2 = \|P_V(w) - v\|^2 + \|P_{V^\perp}(w)\|^2 \geqslant \|P_{V^\perp}(w)\|^2$ so the minimum is attained if $v = P_V(w)$.

Properties (continued)

Proof.

ii) recall $w = P_V(w) + P_{V^\perp}(w)$, then for any $v \in V$, by the Pythagorean theorem $\|w - v\|^2 = \|(P_V(w) - v) + P_{V^\perp}(w)\|^2 = \|P_V(w) - v\|^2 + \|P_{V^\perp}(w)\|^2 \geqslant \|P_{V^\perp}(w)\|^2$ so the minimum is attained if $v = P_V(w)$.

iii) $w - (\frac{w \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{w \cdot v_2}{v_2 \cdot v_2} v_2 + \ldots + \frac{w \cdot v_k}{v_k \cdot v_k} v_k) \in V^\perp$. $\square$

Properties (continued)

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then

i) $P_V \circ P_V = P_V$,

Properties (continued)

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then

i) $P_V \circ P_V = P_V$,

ii) $S_V \circ S_V = \mathrm{id}_{\mathbb{R}^n}$,

Properties (continued)

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then

i) $P_V \circ P_V = P_V$,

ii) $S_V \circ S_V = \mathrm{id}_{\mathbb{R}^n}$,

iii) $P_V + P_{V^\perp} = \mathrm{id}_{\mathbb{R}^n}$,

Properties (continued)

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then

i) $P_V \circ P_V = P_V$,

ii) $S_V \circ S_V = \mathrm{id}_{\mathbb{R}^n}$,

iii) $P_V + P_{V^\perp} = \mathrm{id}_{\mathbb{R}^n}$,

iv) $S_V = -S_{V^\perp}$.

Properties (continued)

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace. Then

i) $P_V \circ P_V = P_V$,

ii) $S_V \circ S_V = \mathrm{id}_{\mathbb{R}^n}$,

iii) $P_V + P_{V^\perp} = \mathrm{id}_{\mathbb{R}^n}$,

iv) $S_V = -S_{V^\perp}$.

Example

Let $V = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 - x_2 + 2x_3 - 2x_4 = 0\}$ and $w = (1, 0, 1, -1)$. Compute $P_V(w)$. By definition $V^\perp = \mathrm{lin}((1, -1, 2, -2))$. Then $P_{V^\perp}(w) = \frac{w \cdot (1,-1,2,-2)}{1^2+(-1)^2+2^2+(-2)^2}(1, -1, 2, -2) = \frac{1}{2}(1, -1, 2, -2)$. Hence $P_V(w) = w - P_{V^\perp}(w) = (\frac{1}{2}, \frac{1}{2}, 0, 0)$.

Gram-Schmidt process

Let $v_1, \ldots, v_k$ be a basis of the subspace $V \subset \mathbb{R}^n$. **The Gram-Schmidt process** is an inductive way of computing an orthogonal basis $w_1, \ldots, w_k$ of V.

Gram-Schmidt process

Let $v_1, \ldots, v_k$ be a basis of the subspace $V \subset \mathbb{R}^n$. **The Gram-Schmidt process** is an inductive way of computing an orthogonal basis $w_1, \ldots, w_k$ of V.
By induction

i) for $i = 1$ set

$$w_1 = v_1, \quad W_1 = \mathrm{lin}(w_1),$$

Gram-Schmidt process

Let $v_1, \ldots, v_k$ be a basis of the subspace $V \subset \mathbb{R}^n$. **The Gram-Schmidt process** is an inductive way of computing an orthogonal basis $w_1, \ldots, w_k$ of V.
By induction

i) for $i = 1$ set

$$w_1 = v_1, \quad W_1 = \mathrm{lin}(w_1),$$

ii) for $1 < i \leqslant k$ set

$$w_i = v_i - P_{W_{i-1}}(v_i),$$

$$W_i = \mathrm{lin}(w_1, \ldots, w_i).$$

Gram-Schmidt process (continued)

Proposition (Gram-Schmidt)

With notation as above for $i = 1, \ldots, k$

i) $w_1, \ldots, w_i$ *is an orthogonal basis of* W_i,

Gram-Schmidt process (continued)

Proposition (Gram-Schmidt)

With notation as above for $i = 1, \ldots, k$

i) $w_1, \ldots, w_i$ *is an orthogonal basis of* W_i,

ii) $W_i = \mathrm{lin}(v_1, \ldots, v_i)$.

Gram-Schmidt process (continued)

Proposition (Gram-Schmidt)

With notation as above for $i = 1, \ldots, k$

i) $w_1, \ldots, w_i$ *is an orthogonal basis of* W_i,

ii) $W_i = \mathrm{lin}(v_1, \ldots, v_i)$.

Gram-Schmidt process (continued)

Proposition (Gram-Schmidt)

With notation as above for $i = 1, \ldots, k$

i) $w_1, \ldots, w_i$ *is an orthogonal basis of* W_i,

ii) $W_i = \mathrm{lin}(v_1, \ldots, v_i)$.

Since $W_k = V$ vectors $w_1, \ldots, w_k$ form an orthogonal basis of V. The normalized vectors $\frac{w_1}{\|w_1\|}, \ldots, \frac{w_k}{\|w_k\|}$ form an orthonormal basis of V.

Example

Let $V = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 - x_2 + 2x_3 - x_4 = 0\}$ and let $v_1 = (1, 0, 0, 1), v_2 = (1, 1, 0, 0), v_3 = (0, 1, 1, 1) \in \mathbb{R}^4$ be a basis of subspace V. Then $w_1 = v_1,\ W_1 = \mathrm{lin}(w_1)$,
$w_1 = (1, 0, 0, 1)$,

Example

Let $V = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 - x_2 + 2x_3 - x_4 = 0\}$ and let $v_1 = (1, 0, 0, 1), v_2 = (1, 1, 0, 0), v_3 = (0, 1, 1, 1) \in \mathbb{R}^4$ be a basis of subspace V. Then $w_1 = v_1,\ W_1 = \mathrm{lin}(w_1)$,

$w_1 = (1, 0, 0, 1)$,

$w_2 = v_2 - P_{W_1}(v_2) = v_2 - \frac{v_2 \cdot w_1}{w_1 \cdot w_1} w_1$,

$w_2 = (1, 1, 0, 0) - \frac{1}{2}(1, 0, 0, 1) = \frac{1}{2}(1, 2, 0, -1),\ W_2 = \mathrm{lin}(w_1, w_2)$

Example

Let $V = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 - x_2 + 2x_3 - x_4 = 0\}$ and let $v_1 = (1, 0, 0, 1), v_2 = (1, 1, 0, 0), v_3 = (0, 1, 1, 1) \in \mathbb{R}^4$ be a basis of subspace V. Then $w_1 = v_1,\ W_1 = \mathrm{lin}(w_1)$,

$w_1 = (1, 0, 0, 1)$,

$w_2 = v_2 - P_{W_1}(v_2) = v_2 - \frac{v_2 \cdot w_1}{w_1 \cdot w_1} w_1$,

$w_2 = (1, 1, 0, 0) - \frac{1}{2}(1, 0, 0, 1) = \frac{1}{2}(1, 2, 0, -1),\ W_2 = \mathrm{lin}(w_1, w_2)$

$w_3 = v_3 - P_{W_2}(v_3) = v_3 - \frac{v_3 \cdot w_1}{w_1 \cdot w_1} w_1 - \frac{v_3 \cdot w_2}{w_2 \cdot w_2} w_2$,

$w_3 = (0, 1, 1, 1) - \frac{1}{2}(1, 0, 0, 1) - \frac{1}{6}(1, 2, 0, -1) =$
$\frac{1}{3}(-2, 2, 3, 2), W_3 = \mathrm{lin}(w_1, w_2, w_3)$.

Example

Let $V = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 - x_2 + 2x_3 - x_4 = 0\}$ and let $v_1 = (1,0,0,1), v_2 = (1,1,0,0), v_3 = (0,1,1,1) \in \mathbb{R}^4$ be a basis of subspace V. Then $w_1 = v_1,\ W_1 = \mathrm{lin}(w_1)$,
$w_1 = (1,0,0,1)$,
$w_2 = v_2 - P_{W_1}(v_2) = v_2 - \frac{v_2 \cdot w_1}{w_1 \cdot w_1} w_1$,
$w_2 = (1,1,0,0) - \frac{1}{2}(1,0,0,1) = \frac{1}{2}(1,2,0,-1),\ W_2 = \mathrm{lin}(w_1, w_2)$
$w_3 = v_3 - P_{W_2}(v_3) = v_3 - \frac{v_3 \cdot w_1}{w_1 \cdot w_1} w_1 - \frac{v_3 \cdot w_2}{w_2 \cdot w_2} w_2$,
$w_3 = (0,1,1,1) - \frac{1}{2}(1,0,0,1) - \frac{1}{6}(1,2,0,-1) =$
$\frac{1}{3}(-2,2,3,2), W_3 = \mathrm{lin}(w_1, w_2, w_3)$.
Therefore $(1,0,0,1), (1,2,0,-1), (-2,2,3,2)$ is an orthogonal basis of $V = \mathrm{lin}(v_1, v_2, v_3)$. Moreover $\frac{1}{\sqrt{2}}(1,0,0,1), \frac{1}{\sqrt{6}}(1,2,0,-1), \frac{1}{\sqrt{21}}(-2,2,3,2)$ is an orthonormal basis of V.

Remark

Note that $\frac{w \cdot v}{v \cdot v} v = \frac{w \cdot (\alpha v)}{(\alpha v) \cdot (\alpha v)} (\alpha v)$.

Projection Matrix

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace and let $w_1, \ldots, w_k \in V$ be an orthonormal basis of V. If $Q \in M(n \times k; \mathbb{R})$ is a matrix with columns equal to $w_1, \ldots, w_k$ then

$$M(P_V)_{st}^{st} = QQ^{\intercal}.$$

Projection Matrix

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace and let $w_1, \ldots, w_k \in V$ be an orthonormal basis of V. If $Q \in M(n \times k; \mathbb{R})$ is a matrix with columns equal to $w_1, \ldots, w_k$ then

$$M(P_V)_{st}^{st} = QQ^{\intercal}.$$

Proof.

Follows from the formula

$$P_V(v) = (v \cdot w_1)w_1 + (v \cdot w_2)w_2 + \ldots + (v \cdot w_k)w_k.$$

$\square$

Projection Matrix

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace and let $w_1, \ldots, w_k \in V$ be an orthonormal basis of V. If $Q \in M(n \times k; \mathbb{R})$ is a matrix with columns equal to $w_1, \ldots, w_k$ then

$$M(P_V)_{st}^{st} = QQ^{\mathsf{T}}.$$

Proof.

Follows from the formula

$$P_V(v) = (v \cdot w_1)w_1 + (v \cdot w_2)w_2 + \ldots + (v \cdot w_k)w_k.$$

$\square$

Remark

Note that $Q^{\mathsf{T}}Q = I_k$.

QR Decomposition

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace and let $v_1, \ldots, v_k \in V$ be a basis of V. Then there exists an orthonormal basis $w_1, \ldots, w_k \in V$ of V and an upper triangular matrix $R \in M(k \times k; \mathbb{R})$ with with positive entries on the diagonal (hence invertible), such that if $A \in M(n \times k; \mathbb{R})$ is a matrix with columns $v_1, \ldots, v_k$ and $Q \in M(n \times k; \mathbb{R})$ is a matrix with columns $w_1, \ldots, w_k$ then

$$A = QR.$$

Example

In the previous example for

$$v_1 = (1,0,0,1), v_2 = (1,1,0,0), v_3 = (0,1,1,1),$$

by the Gram-Schmidt process we have

$$\begin{aligned} w_1 &= v_1, \\ w_2 &= v_2 - \frac{1}{2} w_1, \\ w_3 &= v_3 - \frac{1}{2} w_1 - \frac{1}{3} w_2. \end{aligned}$$

Example

In the previous example for

$$v_1 = (1,0,0,1), v_2 = (1,1,0,0), v_3 = (0,1,1,1),$$

by the Gram-Schmidt process we have

$$\begin{aligned} w_1 &= v_1, \\ w_2 &= v_2 - \frac{1}{2} w_1, \\ w_3 &= v_3 - \frac{1}{2} w_1 - \frac{1}{3} w_2. \end{aligned}$$

Therefore

$$\begin{aligned} v_1 &= w_1, \\ v_2 &= \frac{1}{2} w_1 + w_2, \\ v_3 &= \frac{1}{2} w_1 + \frac{1}{3} w_2 + w_3. \end{aligned}$$

Example (continued)

Since $w_1 = (1,0,0,1), w_2 = \frac{1}{2}(1,2,0,-1), w_3 = \frac{1}{3}(-2,2,3,2)$,

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & -\frac{2}{3} \\ 0 & 1 & \frac{2}{3} \\ 0 & 0 & 1 \\ 1 & -\frac{1}{2} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix}$$

Columns of the first matrix have to be normalized, i.e. divided respectively by $\sqrt{2}, \frac{\sqrt{6}}{2}, \frac{\sqrt{21}}{3}$ (and rows of the second matrix multiplied respectively by the same numbers), hence

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{21}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{2}{\sqrt{21}} \\ 0 & 0 & \frac{3}{\sqrt{21}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{21}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{6}}{2} & \frac{\sqrt{6}}{6} \\ 0 & 0 & \frac{\sqrt{21}}{3} \end{bmatrix}$$

QR Decomposition (continued)

Proof.

Follows directly form the Gram–Schmidt process. In fact

$$R = \begin{bmatrix} \|w_1\| & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|w_k\| \end{bmatrix} \cdot \begin{bmatrix} 1 & \frac{v_2 \cdot w_1}{w_1 \cdot w_1} & \frac{v_3 \cdot w_1}{w_1 \cdot w_1} & \cdots & \frac{v_k \cdot w_1}{w_1 \cdot w_1} \\ 0 & 1 & \frac{v_3 \cdot w_2}{w_2 \cdot w_2} & \cdots & \frac{v_k \cdot w_2}{w_2 \cdot w_2} \\ 0 & 0 & 1 & \cdots & \frac{v_k \cdot w_3}{w_3 \cdot w_3} \\ & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

and

$$Q = \begin{bmatrix} \frac{w_1^\mathsf{T}}{\|w_1\|} \cdots \frac{w_k^\mathsf{T}}{\|w_k\|} \end{bmatrix}.$$

$\square$

Uniqueness of QR Decomposition

Remark

For any matrix $A \in M(n \times k; \mathbb{R})$ *with* $r(A) = k$ *the matrices* Q *and* R *are unique if diagonal entries of* R *are positive. That is if* $A = QR = Q'R'$ *where* $Q, Q' \in M(n \times k; \mathbb{R})$ *are orthogonal matrices and matrices* $R, R' \in M(k \times k; \mathbb{R})$ *are upper triangular matrices with positive entries on the diagonal then* $Q = Q', R = R'$.

Uniqueness of QR Decomposition

Remark

For any matrix $A \in M(n \times k; \mathbb{R})$ with $r(A) = k$ the matrices Q and R are unique if diagonal entries of R are positive. That is if $A = QR = Q'R'$ where $Q, Q' \in M(n \times k; \mathbb{R})$ are orthogonal matrices and matrices $R, R' \in M(k \times k; \mathbb{R})$ are upper triangular matrices with positive entries on the diagonal then $Q = Q', R = R'$.

Proof.

Since $Q'^{\mathsf{T}} Q' = I_k$ from

$$QR = Q'R',$$

by multiplying by Q'^{T} on the left and by R^{-1} on the right we get

$$Q'^{\mathsf{T}} Q = R'R^{-1}.$$

The inverse and product of two upper triangular matrices is upper triangular and the matrix $Q'^{\mathsf{T}} Q$ is orthogonal.

Uniqueness of QR Decomposition

Remark

For any matrix $A \in M(n \times k; \mathbb{R})$ with $r(A) = k$ the matrices Q and R are unique if diagonal entries of R are positive. That is if $A = QR = Q'R'$ where $Q, Q' \in M(n \times k; \mathbb{R})$ are orthogonal matrices and matrices $R, R' \in M(k \times k; \mathbb{R})$ are upper triangular matrices with positive entries on the diagonal then $Q = Q', R = R'$.

Proof.

Since $Q'^{\mathsf{T}} Q' = I_k$ from

$$QR = Q'R',$$

by multiplying by Q'^{T} on the left and by R^{-1} on the right we get

$$Q'^{\mathsf{T}} Q = R'R^{-1}.$$

The inverse and product of two upper triangular matrices is upper triangular and the matrix $Q'^{\mathsf{T}} Q$ is orthogonal. The only upper triangular and orthogonal matrix with positive diagonal entries is the unit matrix.

Projection Matrix (continued)

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace and let $v_1, \ldots, v_k \in V$ be a basis of V. If $A \in M(n \times k; \mathbb{R})$ is a matrix with columns $v_1, \ldots, v_k$ then

$$M(P_V)_{st}^{st} = A(A^\intercal A)^{-1} A^\intercal.$$

Projection Matrix (continued)

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace and let $v_1, \ldots, v_k \in V$ be a basis of V. If $A \in M(n \times k; \mathbb{R})$ is a matrix with columns $v_1, \ldots, v_k$ then

$$M(P_V)_{st}^{st} = A(A^\intercal A)^{-1} A^\intercal.$$

In particular, the matrix $A^\intercal A$ is invertible.

Projection Matrix (continued)

Proposition

Let $V \subset \mathbb{R}^n$ be a subspace and let $v_1, \ldots, v_k \in V$ be a basis of V. If $A \in M(n \times k; \mathbb{R})$ is a matrix with columns $v_1, \ldots, v_k$ then

$$M(P_V)_{st}^{st} = A(A^\intercal A)^{-1} A^\intercal.$$

In particular, the matrix $A^\intercal A$ is invertible.

Proof.

By the QR decomposition there exist matrices $Q \in M(n \times k; \mathbb{R})$ and $R \in M(k \times k; \mathbb{R})$ such that $A = QR$ and

$$M(P_V)_{st}^{st} = QQ^\intercal$$

.

Projection Matrix (continued)

Proof.

Since $A = QR$ the matrix

$$A^\intercal A = (R^\intercal Q^\intercal)(QR) = R^\intercal R,$$

is invertible. Moreover

$$Q = AR^{-1}, \quad Q^\intercal = (R^{-1})^\intercal A^\intercal,$$

Projection Matrix (continued)

Proof.

Since $A = QR$ the matrix

$$A^\intercal A = (R^\intercal Q^\intercal)(QR) = R^\intercal R,$$

is invertible. Moreover

$$Q = AR^{-1}, \quad Q^\intercal = (R^{-1})^\intercal A^\intercal,$$

and

$$M(P_V)_{st}^{st} = QQ^\intercal = (AR^{-1})\left((R^{-1})^\intercal A^\intercal\right) = A(R^\intercal R)^{-1}A^\intercal =$$

$$= A(A^\intercal A)^{-1}A^\intercal.$$

$\square$

Example

Let $V = \mathrm{lin}((1,1,1),(1,2,0))$. Find the formula of the orthogonal projection onto V and the formula of the orthogonal reflection across V.

Example

Let $V = \mathrm{lin}((1,1,1),(1,2,0))$. Find the formula of the orthogonal projection onto V and the formula of the orthogonal reflection across V.

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix}$, so $A^\intercal A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$ and $(A^\intercal A)^{-1} = \frac{1}{6}\begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix}$.

Therefore

$$M(P_V)_{st}^{st} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix} \frac{1}{6}\begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 2 & 2 & 2 \\ 2 & 5 & -1 \\ 2 & -1 & 5 \end{bmatrix}.$$

$$P_V((x_1,x_2,x_3)) = \left(\frac{1}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3, \frac{1}{3}x_1 + \frac{5}{6}x_2 - \frac{1}{6}x_3, \frac{1}{3}x_1 - \frac{1}{6}x_2 + \frac{5}{6}x_3\right).$$

Example (continued)

Moreover,

$$M(S_V)_{st}^{st} = 2M(P_V)_{st}^{st} - I = \frac{1}{3}\begin{bmatrix} 2 & 2 & 2 \\ 2 & 5 & -1 \\ 2 & -1 & 5 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} =$$

$$= \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & 2 & -1 \\ 2 & -1 & 2 \end{bmatrix}$$

and

$$S_V((x_1, x_2, x_3)) = \left(-\frac{1}{3}x_1 + \frac{2}{3}x_2 + \frac{2}{3}x_3, \frac{2}{3}x_1 + \frac{2}{3}x_2 - \frac{1}{3}x_3,\right.$$

$$\left.\frac{2}{3}x_1 - \frac{1}{3}x_2 + \frac{2}{3}x_3\right).$$

Projection Matrix (continued)

Corollary

If P_V is an orthogonal projection and S_V orthogonal reflection then the matrices $M(P_V)_{st}^{st}, M(S_V)_{st}^{st}$ are symmetric.

Projection Matrix (continued)

Corollary

If P_V is an orthogonal projection and S_V orthogonal reflection then the matrices $M(P_V)_{st}^{st}, M(S_V)_{st}^{st}$ are symmetric.

Proposition

If matrix A has linearly independent columns then $A^\top A$ is positive definite. If A is any matrix then $A^\top A$ is positive semidefinite.

Projection Matrix (continued)

Corollary

If P_V is an orthogonal projection and S_V orthogonal reflection then the matrices $M(P_V)_{st}^{st}, M(S_V)_{st}^{st}$ are symmetric.

Proposition

If matrix A has linearly independent columns then $A^\intercal A$ is positive definite. If A is any matrix then $A^\intercal A$ is positive semidefinite.

Proof.

For any vector $v \in \mathbb{R}^n$ we have $\|Av\|^2 = v^\intercal A^\intercal Av \geqslant 0$. When A has linearly independent columns then $Av = 0$ implies that $v = 0$. $\square$

Projection Matrix (continued)

Proposition

Matrix representing an oblique projection onto $\operatorname{im}(A)$ *along* $\ker B$ *for any matrices* $A \in M(n \times k; \mathbb{R})$, $B \in M(l \times n; \mathbb{R})$ *such that* $\operatorname{im}(A)$ *and* $\ker B$ *give a direct sum decomposition is given by*

$$P = A(B^\intercal A)^+ B^\intercal.$$

Proof.

(sketch, cf. Lecture 14). The assumptions imply that $\operatorname{rank} A = \operatorname{rank} B = \operatorname{rank}(B^\intercal A)$ as $\operatorname{im}(A) \cap \ker B = \{0\}$. It is well–known that with such assumptions the inverse law holds, i.e.,

$$(B^\intercal A)^+ = A^+ (B^\intercal)^+.$$

Finally

$$P = \left(AA^+\right)\left((B^\intercal)^+ B^\intercal\right) = \left(AA^+\right)\left(BB^+\right),$$

and clearly $P^2 = P$ with $\operatorname{im}(P) = \operatorname{im}(A)$ and $\ker P = \ker B$. $\square$

Projection onto Intersection of Subspaces

Proposition

Let $A, B \in M(n \times n; \mathbb{R})$ be symmetric and positive semidefinite matrices. Then

$$\operatorname{im}(A + B) = \operatorname{im}(A) + \operatorname{im}(B).$$

Proof.

Obviously $\operatorname{im}(A + B) \subset \operatorname{im}(A) + \operatorname{im}(B)$. It is enough to prove that $\operatorname{im}(A) \subset \operatorname{im}(A + B)$ which is equivalent to

$$\ker(A + B) \subset \ker(A).$$

Let $(A + B)x = 0$. Then $x^\intercal AX + x^\intercal Bx = 0$. Matrices A, B are positive semidefinite hence $x^\intercal Ax = 0$ which is equivalent to $Ax = 0$. $\square$

Projection onto Intersection of Subspaces (continued)

Lemma (Anderson,Duffin)

Let $P = AA^+$ *and* $Q = BB^+$ *be matrices of orthogonal projections onto* $\mathrm{im}(A)$ *and* $\mathrm{im}(B)$ *with respect to the standard basis, respectively. Let* $V = \mathrm{im}(A) \cap \mathrm{im}(B)$*. Then*

$$M(P_V)_{st}^{st} = 2P(P+Q)^+Q.$$

Proof.

In the proof we use repeatedly the fact that the for subspaces $V, W \subset \mathbb{R}^n$

$$(V+W)^\perp = V^\perp \cap W^\perp.$$

Let

$$T = M(P_V)_{st}^{st}.$$

Projection onto Intersection of Subspaces (continued)

Proof.

Matrix T does not depend on the order of P and Q as

$$P(P+Q)^{+}Q = (P+Q-Q)(P+Q)^{+}(P+Q-P) =$$

$$= (P+Q)(P+Q)^{+}(P+Q) - Q(P+Q)^{+}(P+Q) -$$

$$-(P+Q)(P+Q)^{+}P + Q(P+Q)^{+}P =$$

$$= (I-(P+Q)(P+Q)^{+})P + Q(I-(P+Q)^{+}(P+Q)) + Q(P+Q)^{+}P =$$

$$= Q(P+Q)^{+}P.$$

First two terms are zero, $(P+Q)(P+Q)^{+}$ is a matrix of the orthogonal projection onto $\mathrm{im}(P)+\mathrm{im}(Q)$, similarly $(P+Q)^{+}(P+Q)$ is a matrix of the orthogonal projection onto $\mathrm{im}((P+Q)^{\mathsf{T}}) = \mathrm{im}(P)+\mathrm{im}(Q)$ with kernel $(\mathrm{im}(P)+\mathrm{im}(Q))^{\mathsf{T}} = \ker(P)\cap\ker(Q)$.

Projection onto Intersection of Subspaces (continued)

Proof.

Therefore

$$\mathrm{im}(T) \subset \mathrm{im}(P) \cap \mathrm{im}(Q) = \mathrm{im}(A) \cap \mathrm{im}(B).$$

On the other hand let $x \in \mathrm{im}(P) \cap \mathrm{im}(Q)$. Then (note that $P^+ = P$ and $Px = x$, similarly for Q)

$$2P(P+Q)^+Qx = P(P+Q)^+Q\big(P^+ + Q^+\big)x =$$

$$= P(P+Q)^+QQ^+x + Q(P+Q)^+PP^+x =$$

$$= P(P+Q)^+x + Q(P+Q)^+x = (P+Q)(P+Q)^+x = x.$$

Since P, Q are symmetric T is symmetric too. By the above equation matrix T acts as identity on the subspace $\mathrm{im}(P) \cap \mathrm{im}(Q)$ and it zero on its orthogonal complement, i.e., $\ker(P) + \ker(Q)$. Therefore $T^2 = T$, i.e. T is a matrix of orthogonal projection. $\square$

Linear Least Squares

Let $AX = B$ be a system of n linear equations in k variables with $A \in M(n \times k; \mathbb{R})$ such that $r(A) = k$ (therefore $n \geqslant k$).

Linear Least Squares

Let $AX = B$ be a system of n linear equations in k variables with $A \in M(n \times k; \mathbb{R})$ such that $r(A) = k$ (therefore $n \geqslant k$). If $r(A) = r([A|B])$ the system has a unique solution. Otherwise one can ask what is **the best–fitting solution**, i.e. for X with $\|AX - B\|$ minimal.

Linear Least Squares

Let $AX = B$ be a system of n linear equations in k variables with $A \in M(n \times k; \mathbb{R})$ such that $r(A) = k$ (therefore $n \geqslant k$). If $r(A) = r([A|B])$ the system has a unique solution. Otherwise one can ask what is **the best–fitting solution**, i.e. for X with $\|AX - B\|$ minimal. The least value is the distance of vector B from the subspace $V \subset \mathbb{R}^n$ spanned by the columns of A and it is realized by the orthogonal projection of B on V.

Linear Least Squares

Let $AX = B$ be a system of n linear equations in k variables with $A \in M(n \times k; \mathbb{R})$ such that $r(A) = k$ (therefore $n \geqslant k$). If $r(A) = r([A|B])$ the system has a unique solution. Otherwise one can ask what is **the best–fitting solution**, i.e. for X with $\|AX - B\|$ minimal. The least value is the distance of vector B from the subspace $V \subset \mathbb{R}^n$ spanned by the columns of A and it is realized by the orthogonal projection of B on V.

By the previous formula

$$P_V(B) = A(A^\intercal A)^{-1} A^\intercal B.$$

The unique solution of the equation $AX = A(A^\intercal A)^{-1} A^\intercal B$ is

$$X = (A^\intercal A)^{-1} A^\intercal B.$$

Linear Least Squares (continued)

The name (linear) last squares comes from the fact we look for $(x_1, \ldots, x_k) \in \mathbb{R}^k$ for which the value of

$$
\begin{aligned}
&(a_{11}x_1 + a_{12}x_2 + \ldots + a_{1k}x_k - b_1)^2 + \\
&+ (a_{21}x_1 + a_{22}x_2 + \ldots + a_{2k}x_k - b_2)^2 + \\
&\vdots \\
&+ (a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nk}x_k - b_n)^2 = \\
&= \|AX - B\|^2,
\end{aligned}
$$

is minimal.

Linear Least Squares (continued)

The name (linear) last squares comes from the fact we look for $(x_1, \ldots, x_k) \in \mathbb{R}^k$ for which the value of

$$(a_{11}x_1 + a_{12}x_2 + \ldots + a_{1k}x_k - b_1)^2 +$$

$$+ (a_{21}x_1 + a_{22}x_2 + \ldots + a_{2k}x_k - b_2)^2 +$$

$$\vdots$$

$$+ (a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nk}x_k - b_n)^2 =$$

$$= \|AX - B\|^2,$$

is minimal.

Remark

If columns of matrix A are not linearly independent any vector $X \in \mathbb{R}^k$ such that $AX = P_V(B)$, where $V = \mathrm{colsp}(A)$ gives the least $\|AX - B\|$ (such X is not unique).

Example

Find the best–fitting solution of the (inconsistent) system of linear equations.

$$\begin{cases} x_1 + x_2 = 3 \\ 2x_1 + 3x_2 = 0 \\ x_1 + x_2 = -1 \end{cases}$$

Let

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}.$$

Then

$$A^\intercal A = \begin{bmatrix} 6 & 8 \\ 8 & 11 \end{bmatrix}, \quad (A^\intercal A)^{-1} = \begin{bmatrix} \frac{11}{2} & -4 \\ -4 & 3 \end{bmatrix}.$$

The best–fitting solution is equal to

$$X = (A^\intercal A)^{-1} A^\intercal B = \begin{bmatrix} \frac{11}{2} & -4 \\ -4 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix},$$

i.e., $x_1 = 3,\ x_2 = -2$.

Cauchy–Schwarz Inequality and Gram Matrix

Proposition

For any $v, w \in \mathbb{R}^n$

$$\|v\|\|w\| \geqslant |v \cdot w|,$$

and the equality holds if and only if v, w *are linearly dependent.*

Cauchy–Schwarz Inequality and Gram Matrix

Proposition

For any $v, w \in \mathbb{R}^n$

$$\|v\|\|w\| \geqslant |v \cdot w|,$$

and the equality holds if and only if v, w *are linearly dependent.*

Proof.

Let $v = (x_1, \ldots, x_n), w = (y_1, \ldots, y_n)$ and

$$A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{bmatrix}.$$

Cauchy–Schwarz Inequality (continued)

Proof.

By the Cauch–Binet formula

$$\det(AA^{\mathsf{T}}) = \det \begin{bmatrix} \|v\|^2 & v \cdot w \\ w \cdot v & \|w\|^2 \end{bmatrix} =$$

$$= \|v\|^2 \|w\|^2 - (v \cdot w)^2 = \sum_{\substack{J \subset \{1,\ldots,n\} \\ \#J = 2}} (\det A_{2;J})^2 \geqslant 0,$$

and the equality holds if and only if all order 2 minors of A vanish, i.e. the rank of A is either 0 or 1. $\square$

Minkowski Inequality

Proposition

For any $v, w \in \mathbb{R}^n$

$$\|v + w\| \leqslant \|v\| + \|w\|,$$

and the equality holds if and only if there exist $\alpha \geqslant 0$ *such that* $v = \alpha w$ *or* $w = 0$.

Proof.

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2(v \cdot w) \leqslant \|v\|^2 + \|w\|^2 + 2\|v\|\|w\| =$$

$$= (\|v\| + \|w\|)^2,$$

and the equality holds if and only if $v \cdot w = \|v\|\|w\|$, i.e v and w are linearly dependent. If $v, w \neq 0$ this happens if and only if $\cos \angle(v, w) = 1$, that is $\angle(v, w) = 0$. $\square$

Covariance as Scalar Product

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let

$$V = \{X \colon \Omega \to \mathbb{R} \mid X \text{ is measurable, } \mathrm{E}(X) = 0, \ \mathrm{Var}(X) < \infty\}/\sim,$$

where for $X, Y \colon \Omega \to \mathbb{R}$

$$X \sim Y \iff P(\{\omega \in \Omega \mid X(\omega) = Y(\omega)\}) = 1,$$

(i.e. we identify random variables X and Y if they are equal with probability 1). Then V is a vector space and

$$X \cdot Y = \mathrm{Cov}(X, Y) = \mathrm{E}(XY),$$

is a scalar product. Under this definition $\|X\|$ is the standard deviation of X, X is perpendicular to Y if and only if X and Y are uncorrelated, and

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}},$$

is the cosine of the angle between X and Y.

Centering Matrix

Let

$$\mathbb{1} = (1, \ldots, 1).$$

For any vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ define

$$\overline{\boldsymbol{x}} = \frac{1}{n}(x_1 + \ldots + x_n)\mathbb{1} = \frac{1}{n}\mathbb{1}^\intercal \boldsymbol{x}\mathbb{1}.$$

The **centered** vector $\boldsymbol{x}$ is

$$\boldsymbol{x} - \overline{\boldsymbol{x}}.$$

Observe that

$$\overline{\boldsymbol{x} - \overline{\boldsymbol{x}}} = 0.$$

The centering matrix is

$$H = I - \frac{1}{n}\mathbb{1}\mathbb{1}^\intercal.$$

Sample Pearson Correlation Coefficient

Let

$$\boldsymbol{x} = (x_1, \ldots, x_n),$$

$$\boldsymbol{y} = (y_1, \ldots, y_n).$$

The sample Pearson correlation coefficient is

$$r_{xy} = \frac{(\boldsymbol{x} - \overline{\boldsymbol{x}}) \cdot (\boldsymbol{y} - \overline{\boldsymbol{y}})}{\|\boldsymbol{x} - \overline{\boldsymbol{x}}\| \|\boldsymbol{y} - \overline{\boldsymbol{y}}\|},$$

that is a cosine between centered vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Therefore

$$-1 \leqslant r_{xy} \leqslant 1.$$

It is a measure of similarity between data of (centered) $\boldsymbol{x}$ and $\boldsymbol{y}$.

Sample Covariance Matrix

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^m$. Let

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix},$$

be a matrix whose columns consists of vectors $\boldsymbol{x}_i$. The **sample covariance matrix** $K = [k_{ij}] \in M(n \times n; \mathbb{R})$ such that

$$k_{ij} = \frac{1}{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}_i}) \cdot (\boldsymbol{x}_j - \overline{\boldsymbol{x}_j}).$$

It is a symmetric matrix, given by the formula

$$K = \frac{1}{n}\boldsymbol{X}^\intercal H \boldsymbol{X}.$$

Sample Covariance Matrix (continued)

The matrix

$$\overline{\boldsymbol{X}} = \begin{bmatrix} \overline{\boldsymbol{x}}_1 & \cdots & \overline{\boldsymbol{x}}_n \end{bmatrix},$$

with centered columns is given by the formula

$$\overline{\boldsymbol{X}} = \boldsymbol{X} - \frac{1}{n}\mathbb{1}\mathbb{1}^\intercal \boldsymbol{X} = H\boldsymbol{X}.$$

Therefore

$$K = \frac{1}{n}\overline{\boldsymbol{X}}^\intercal \overline{\boldsymbol{X}} = \frac{1}{n}(H\boldsymbol{X})^\intercal (H\boldsymbol{X}) =$$

$$= \frac{1}{n}\boldsymbol{X}^\intercal H^\intercal H\boldsymbol{X} = \frac{1}{n}\boldsymbol{X}^\intercal H\boldsymbol{X},$$

as $H^\intercal = H$ and $H^2 = H$. It follows that the sample covariance matrix is positive semidefinite.

Sample Correlation Matrix

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix},$$

be a matrix whose columns consists of vectors $\boldsymbol{x}_i$. The **sample correlation matrix** $R = [r_{ij}] \in M(n \times n; \mathbb{R})$ such that

$$r_{ij} = \frac{(\boldsymbol{x}_i - \overline{\boldsymbol{x}_i}) \cdot (\boldsymbol{x}_j - \overline{\boldsymbol{x}_j})}{\|\boldsymbol{x}_i - \overline{\boldsymbol{x}_i}\| \|\boldsymbol{x}_j - \overline{\boldsymbol{x}_j}\|} = \frac{k_{ij}}{\sqrt{k_{ii}}\sqrt{k_{jj}}}.$$

It is a symmetric matrix containing all sample Pearson correlation coefficient of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Let

$$D = \operatorname{diag}(k_1, \ldots, k_n),$$

where $k_i = \sqrt{k_{ii}}$. Then

$$K = D^\intercal R D,$$

or equivalently

$$R = (D^{-1})^\intercal K D^{-1}.$$